HAL
open science

The price of connectedness in graph partitioning problems

Nicolas Martin, Paolo Frasca, Takayuki Ishizaki, Jun-Ichi Imura, Carlos Canudas de Wit

To cite this version:

Nicolas Martin, Paolo Frasca, Takayuki Ishizaki, Jun-Ichi Imura, Carlos Canudas de Wit. The price of connectedness in graph partitioning problems. ECC 2019 - European Control Conference, Jun 2019, Naples, Italy. hal-02171099v1

HAL Id: hal-02171099

https://hal.science/hal-02171099v1

Submitted on 2 Jul 2019 (v1), last revised 19 Aug 2019 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The price of connectedness in graph partitioning problems

Nicolas Martin[1], Paolo Frasca[1], Takayuki Ishizaki[2], Jun-Ichi Imura[2], Carlos Canudas-de-Wit[3]

***Abstract*— In this work, we present some results coming from the comparison of two classes of problems of optimal graph partitioning. In the first class of problem, called further *constrained problem*, ones impose the nodes belonging to a same cluster to be connected. In the second class of problem, no constraint of connectedness is imposed. Obviously, the optimal solution of an unconstrained problem is always better than or equal to the solution of the corresponding constrained problem. It is thus interesting to estimate how much the connectedness constraint degrades the optimal solution of the partitioning problem. This degradation is what we call here the *price of connectedness*. Motivated by a numerical example, we propose to estimate the price of connectedness by comparing the cardinalities of the feasible sets in the unconstrained and constrained optimisation problems. We present then a tight upper-bound on the ratio of these two cardinalities for a directed Erdős-Rényi graph. On our way to set this result we also derive an upper-bound on the probability that an Erdős-Rényi graph is connected and the exact probability that an Erdős-Rényi graph contains isolated nodes.**

I. INTRODUCTION

Network partitioning aims to obtain a reduced network abstracting a large-scale network by merging group of nodes into super-nodes. The objective of network partition methods is to cut the complexity of the original network while preserving some properties. Most problems of partition can be written as optimisation problems where one seek the best partition under some constraints. In this article we consider this kind of optimal graph partition problems. Following the application, the complexity considered and the properties to preserve may vary and a large number of works have treated this topic. For example, [1] aims to preserve stability and synchronisation of the system, [2] preserves the network structure, [3] preserves the flow network property and the eigenvector centrality and [4] provides a reduced system with a dynamical behaviour close to the initial system while preserving several properties for control purpose. This last work will be presented in this article as a motivating example. In the general case, nodes in a same cluster of the partition may be not connected. However, for some applications it is necessary that the nodes belonging to a same cluster are connected together. A partition respecting this last property is referred as a connected partition. This constraint may be necessary, for example, when one considers real network systems having a geographical nature such as traffic road network or power grid [3]. Moreover in some problems the connectedness constraint may or may not be imposed. In these cases it is useful to estimate the price of connectedness. As we consider directed graph, connectedness may have different definitions. We will consider here weak connectedness: a directed graph is said to be weakly connected if the underlying undirected graph (obtained by replacing directed edges by simple edges) is connected. The weak connectedness constraint is enough to ensures the preservation of the geographical nature of the graph.

When the connectedness constraint is considered, it is clear that the optimal solution is always worse or equal to the optimal solution without the connectedness constraint. In this article we investigate how much the connectedness constraint degrades the solution. The difference between the two optimal solutions is what we call *the price of connectedness*, denoted by Δ. The approach proposed here allows to tackle the problem at a high level without looking at the particular partitioning problem. We propose to estimate the price of connectedness with the ratio of the cardinalities of the two research sets: the set of partitions and the set of connected partitions. This ratio is called *ratio of connectedness* and is denoted by ρ. To motivate this approach we will present some simulations on a partition problem from the literature: the cluster model reduction [5]. We focus then on the estimation of the ratio of connectedness ρ. This ratio can also be viewed as the probability that a random partition is a connected partition. We consider in this article random directed graph obtained via the directed Erdős-Rényi model detailed later and unless otherwise mentioned, in the following, a graph stands for a directed graph. We show then that ρ in an Erdős-Rényi graph is directly linked with the probability $\bar{\rho}$ that an Erdős-Rényi graph is connected. The value of this latter probability has been investigated by a broad literature providing asymptotic estimations. See [6] for a discussion on these results. These asymptotic estimations are valid for very large graphs. We propose here an upper bound on $\bar{\rho}$ valid for graphs of any size. From this result we derive an upper bound on the ratio of connectedness ρ. Note that the upper-bound is established for a given size of the clusters. Exactly, this means that for a given directed Erdős-Rényi graph with parameters (n,p) and a vector M where m_i is the number of nodes in the cluster i (such that $\sum_i m_i = n$), we present an upper-bound on $\rho(n,p,M)$. The main contributions of this paper are: i) Emphasis the

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement N694209) www.scale-freeback.eu The collaboration leading to this work has been possible thanks to the Summer Program of the Japan Society for Promotion of Science.

[1] Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France {nicolas.martin;paolo.frasca}@gipsa-lab.fr

[2] Department of Systems and Control Engineering, Graduate School of Engineering, Tokyo Institute of Technology; 2-12-1, Meguro, Tokyo, Japan. {ishizaki, imura}@sc.e.titech.ac.jp

[3] CNRS, GIPSA-lab carlos.canudas-de-wit@gipsa-lab.fr

relation between the ratio of connectedness ρ and the price of connectedness Δ via a numerical example; ii) The link between ρ in an Erdős-Rényi graph and the probability $\bar{\rho}$ that an Erdős-Rényi graph is connected; and finally iii) the derivation of a tight upper-bound on $\bar{\rho}$ leading to a tight upper-bound on ρ, both for Erdős-Rényi graphs. The paper is composed as follows: in the second section an example illustrates the price of connectedness in a concrete problem and the approach that we propose, in the third section, numerical simulations are presented to illustrate the link between ρ and Δ, the fourth section contains the main results, namely the computations leading to the upper-bound on ρ, in a last section we present some conclusions.

II. MOTIVATING EXAMPLE : CLUSTERED MODEL REDUCTION

In this section we will present an example of partitioning problem from the literature. Within this example we will emphasise the price of connectedness which is the gap between the optimal solutions of the constrained and unconstrained problems.

A. The clustered model reduction problem

Consider the following discrete-time linear system:

$$\Sigma : x(t+1) = Ax(t) + Bu(t) \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state of the system. Given $\hat{n}$ the desired size of the reduced system. We define the reduced system as follows:

$$\hat{\Sigma}_P : \begin{cases} \xi(t+1) & = PAP^T\xi(t) + PBu(t) \\ \hat{x}(t) & = P^T\xi(t) \end{cases} \tag{2}$$

where $P \in \mathbb{R}^{\hat{n}\times n}$ is the reduction matrix. The problem is to find the matrix P making the dynamic of $\hat{\Sigma}_P$ the *closest* to the dynamic of Σ. More precisely, we want to minimise $\|g - \hat{g}\|_{\mathscr{H}_2}$, where g is the transfer function from control u to state x and $\hat{g}$ is the transfer function from control u to state $\hat{x}$. A continuous-time version of the problem is presented in [7] and different extensions are presented, for example, in [4], [5] and [8]. The unconstrained problem can be written as follows:

$$\min_{P\in\mathbb{R}^{\hat{n}\times n}} \|g - \hat{g}_P\|_{\mathscr{H}_2}, \tag{3}$$

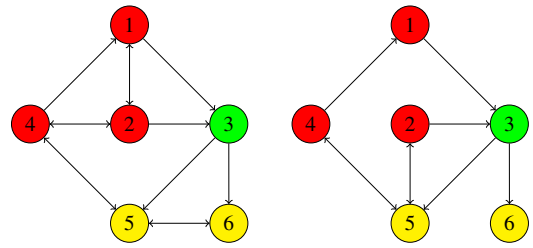
where $\hat{g}_P$ is the transfer function obtained with P in $\hat{\Sigma}$.
To formulate the constrained problem associated to (3) we introduce $I(P)$ the partition associated to the reduction matrix P. We have $I(P) := \{I_1, I_2, \ldots, I_{\hat{n}}\}$ where I_l is a *cluster* of the partition defined as:

$$I_l := \{j \in \{1,\ldots,n\}, P_{l,j} \neq 0\} \tag{4}$$

As an example, let P be the following partition matrix:

$$P = \begin{pmatrix} \star & \star & 0 & \star & 0 & 0 \\ 0 & 0 & \star & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \star & \star \end{pmatrix} \tag{5}$$

where $\star$ represents any non-zero real number, then $I(P) = \{\{1,2,4\},\{3\},\{5,6\}\}$.



(a) With this graph the partition is connected: $I(P) \in \mathbb{I}_c(G)$.

(b) With this graph the partition is not connected: $I(P) \notin \mathbb{I}_c(G)$

Fig. 1: Connected and disconnected partitions. Here we consider the partition defined in eq. 5 applied to two different graphs.

Definition 1: A partition I of a graph G is a connected partition if and only if every subgraph $G_{|I_l}$ induced by a cluster I_l is weakly connected. $\mathbb{I}_c(G)$ represents the set of connected partitions of a graph G.
See Fig. 1 for an example. The constrained problem can be written as follows:

$$\min_{\substack{P\in\mathbb{R}^{\hat{n}\times n} \\ I(P)\in\mathbb{I}_c(G)}} \|g - \hat{g}_P\|_{\mathscr{H}_2} \tag{6}$$

In [5] the authors propose an algorithm to solve problem (3). In appendix A, we present this algorithm and the adaptation of it allowing to solve the constrained problem (6). In the next section, we present some simulations to illustrate the price of connectedness.

B. The price of connectedness

a) Simulations: The Erdős-Rényi model [9] allows to generate random directed graphs as follows: we consider n vertices and each of the n^2 potential edges (self-loop are allowed) exists with a probability p. Here, we consider a collection of 500 Erdős-Rényi graphs with $p = 0.2$ and $n = 100$, and we solve the partition problem without connectedness constraint (3) and with connectedness constraint (6) for each graph. We denote by ε_u the mean error in the unconstrained problem and by ε_c the mean error in the constrained problem. Figure 2 shows the value of ε_u and ε_c in function of the reduction factor $\frac{n-\hat{n}}{n}$.

The relative difference between the two errors represented by an arrow in Fig. 2 is *the price of connectedness*. We denote it by Δ and we define it as:

$$\Delta := \frac{\varepsilon_c - \varepsilon_u}{\varepsilon_u} \tag{7}$$

b) A counting based estimation: Beyond this example, the loss due to the connectedness constraint always exists in any partitioning problem and it is useful to quantify its value. We propose here to tackle this question at a high level: without looking at the particular problem in question we propose to estimate the price of connectedness for any couple of problems which can be formulated as an optimisation problem, which is any couple of problems

$$\min_{I\in\mathbb{I}(G_0)} f(G_0, I) \qquad \text{and} \qquad \min_{I\in\mathbb{I}_c(G_0)} f(G_0, I) \tag{8}$$

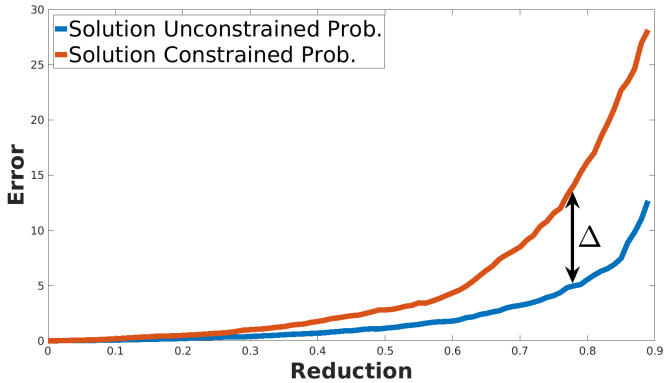


Fig. 2: Estimation error in the unconstrained problem and in the constrained problem. The difference between the two results is the price of connectedness.

which are respectively the Unconstrained Problem and the Constrained Problem. The approach presented here consists in comparing the cardinalities of $\mathbb{I}$ and $\mathbb{I}_c$ which are respectively the set of partition and the set of connected partition of G_0.

More precisely, given a network G having n nodes, and a vector M (verifying $\sum m_i = n$) we want to estimate the ratio of connectedness of defined as follows:

$$\rho(G,M) := \frac{|\mathbb{I}_c(G,M)|}{|\mathbb{I}(G,M)|} \tag{9}$$

where $\mathbb{I}(G,M)$ and $\mathbb{I}_c(G,M)$ are respectively the set of partitions and connected partitions of G verifying $|I_l| = m_l$ (cluster l contains m_l nodes).

Note that in the equation (9), $\mathbb{I}(G,M)$ does not depend on G as the connections between nodes do not matter*.

Before going deeper into the estimation of ρ, we will present in the next section some evidences on the link between the ratio of connectedness ρ and the price of connectedness Δ.

III. Relation between ρ and Δ in the clustered model reduction

A way to emphasise and make more precise the relation between ρ and Δ is to observe simulations on the clustered model reduction presented in the previous section. Precisely, we want to observe how the price of connectedness Δ evolves when the simulations are done with graphs having a different ratio of connectedness ρ. For this purpose we consider 100 Erdős-Rényi graphs with $n = 100$ vertices and an edge probability p varying between 0 and 1†. Figure 3 shows the relation between the ρ and Δ.

By inspection it seems that, in this case, Δ could be approximately related to ρ via the following equation:

$$\Delta(\rho) = -\gamma \log(\rho) \tag{10}$$

*Its cardinality can be computed as in [10]. However we will not use this result since we will derive an upper-bound of ρ and not its exact value.

†It is clear that the ratio of connectedness ρ grows with p: increasing p gives a graph with more edges and so the number of connected partitions is higher. At the limit, if $p = 1$ (complete graph) then $\rho = 1$ and if $p = 0$ (null graph) then $\rho = 0$.

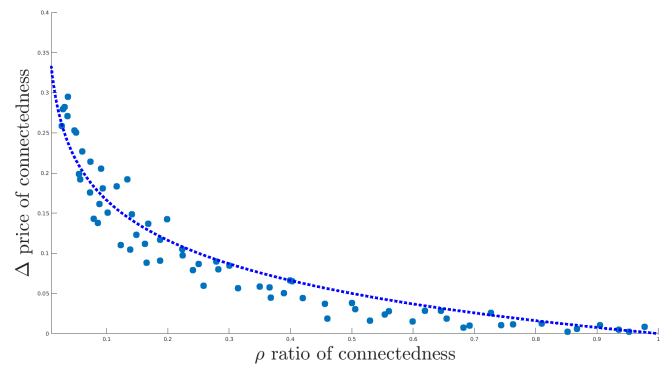


Fig. 3: Relation between the price of connectedness and the ratio of connectedness. Each point correspond to an Erdős-Rényi graph with $n = 100$ and $p \in [0;1]$. The ratio of connectedness, $\rho(G,M)$ is estimated by counting the number of connected partition over the 500 random partition generated (with an arbitrary $M = [5,5,\ldots,5]$). Its price of connectedness $\Delta(G)$ is computed using (7) and Appendix A. In dotted blue the fitting function in Eq.(10).

with $\gamma \approx 0.17$. Moreover, this function verifies $\Delta(\rho) \underset{\rho \to 0}{\longrightarrow} \infty$ and $\Delta(1) = 0$ which is what we expect from the relation between Δ and ρ‡.

While the relation (10) is intrinsically linked to the particular couple of problems (3)-(6), it is clear that Δ and ρ are always related. This relation is intuitive when formulated as follows: the smaller the research set of the constrained problem, the larger the gap between the solutions of the constrained and unconstrained problems.

Motivated by this example, we propose, in the next section, a tight upper-bound on ρ in a (n,p)-Erdős-Rényi graph for a given cluster size vector M.

IV. Value of ρ in an (n,p)-Erdős-Rényi graph

We try here to estimate the ratio of connectedness in a directed (n,p)-Erdős-Rényi graph for a given cluster size vector M.

A. Factorisation formula

Noticing that drawing randomly a cluster of size m in a (n,p) Erdős-Rényi-graph is equivalent to generate a (m,p) Erdős-Rényi-graph, we have:

$$\rho(n,p,M) = \prod_{i=1}^{\hat{n}} \bar{\rho}(m_i,p) \tag{11}$$

where $\bar{\rho}(m_i,p)$ is the probability that a (m,p) Erdős-Rényi-graph is connected. We focus now on $\bar{\rho}(m,p)$ the probability of connectedness of a (m,p)-Erdős-Rényi graph. This question has been investigated in several works. Already the seminal paper of Erdős and Rényi [9] described a phase

‡When ρ goes to 0 the number of connected partitions goes to 0 and the optimal solution of the constrained problem worsens which is $\Delta \to \infty$. On the other hand, when $\rho = 1$, every partition is a connected partition (the graph is complete) and the solutions of the constrained and unconstrained problems are the same which is $\Delta = 0$.

transition behaviour for the value of this probability in the case of undirected graph. Several works [11], [12] investigate the phase transition behaviour in Erdős-Rényi directed graph. However only results for strong connectedness are provided while we look here for weak connectedness. In particular, if we denote by $G_{n,p}$ a random (n,p)-Erdős-Rényi graph and $\mathscr{S}$ the set of strongly connected network it is shown that:

$$\begin{cases} \mathbb{P}(G_{n,p} \in \mathscr{S}) \xrightarrow[n\to\infty]{} 1 & \text{if } p > \frac{\ln n}{n} \\ \mathbb{P}(G_{n,p} \in \mathscr{S}) \xrightarrow[n\to\infty]{} 0 & \text{if } p < \frac{\ln n}{n} \end{cases} \tag{12}$$

From our knowledge there is not such a result for weak connectedness. However, it is clear that if such a threshold exists it should be smaller or equal to $\frac{\ln n}{n}$ (because weak connectedness is a milder property than strong connectedness) and we presume that it is equal. Phase transitions results are valid for large-scale graphs. However in the case of graph partitioning the size of the clusters are relatively small (i.e. m_i in Eq. (11) is small) and so these results can not be used. In the next section we propose an upper-bound on the probability of weak connectedness in directed Erdős-Rényi graph that can be used for any graphs.

B. Upper-bound on $\bar{\rho}$

To estimate the probability $\bar{\rho}(m,p)$, we propose to consider an upper-bounding probability $\bar{\rho}_0(m,p)$ which is the probability that in a (m,p)-Erdős-Rényi graph there is no isolated node (node with degree 0). As the set of connected graphs is included in the set of graphs without isolated nodes, we have naturally $\bar{\rho}_0(m,p) > \bar{\rho}(m,p)$. As for the question of connectedness some works have emphasised a phase transition behaviour for the probability of presence of isolated nodes [13], [14]. However these results are only for undirected graphs and are valid for large-scale graphs. The following proposition gives an exact expression for $\bar{\rho}_0$:

Proposition 1: Let $\bar{\rho}_0(m,p)$ be the probability that a directed (m,p)-Erdős-Rényi graph has no isolated nodes. We have

$$\bar{\rho}_0(m,p) = 1 - \sum_{k=0}^{m^2-2} \theta(m,k) p^k (1-p)^{m^2-k} \tag{13}$$

where $\theta(m,k)$ is the number of directed graphs with m nodes and k edges and with (at least) one isolated node.

Proof: Let us denote Γ_m the set of graph with m nodes having isolated nodes and $\bar{\Gamma}_m$ its complementary. Let $G_{m,p}$ be a random Erdős-Rényi graph

$$\bar{\rho}_0(m,p) = \mathbb{P}(G_{m,p} \in \bar{\Gamma}_m) \tag{14}$$

$$= 1 - \mathbb{P}(G_{m,p} \in \Gamma_m) \tag{15}$$

Now if we denote Γ_m^k the set of graphs of size m with isolated nodes and k edges we have:

$$\bar{\rho}_0(m,p) = 1 - \sum_{k=1}^{m^2} \mathbb{P}(G_{m,p} \in \Gamma_m^k) \tag{16}$$

Every graph in Γ_m^k has the same probability to appear which is the probability that a (m,p)-Erdős-Rényi graph has k edges. Hence we have $\mathbb{P}(G_{m,p} \in \Gamma_m^k) = |\Gamma_m^k| p^k (1-p)^{m^2-k}$. Let us denote $\theta(m,k) := |\Gamma_m^k|$. Noting that $\theta(m,m^2) = \theta(m,m^2-1) = 0$ leads to the result. ■

Moreover, the following theorem provides a value of θ.

Theorem 1: Let $\theta(m,k)$ be the number of graphs with m nodes, k edges and (at least) one isolated node. We have

$$\theta(m,k) = \begin{cases} \binom{m^2}{k} & \text{if } k < \frac{m}{2} \\ \sum_{s=1}^{\lfloor m-\sqrt{k} \rfloor} \binom{m}{s} \left[\binom{(m-s)^2}{k} - \theta(m-s,k) \right] & \text{if } k \geq \frac{m}{2} \end{cases} \tag{17}$$

Proof: When k the number of edges of a graph is strictly smaller than $\frac{m}{2}$ it is impossible that the k edges cover the m nodes. Thus, any graph with m nodes and $k < \frac{m}{2}$ edges contains (at least) one isolated node. The number of such graph is equal to the number of way to arrange k edges among the m^2 possible edges. This number is equal to $\binom{m^2}{k}$. Let's denote $\theta_s(n,k)$ the number of graphs with n nodes, k edges and s isolated nodes. We have:

$$\theta(m,k) = \sum_{s=1}^{m} \theta_s(m,k) \tag{18}$$

Let us remark that $s > m - \sqrt{k} \implies \theta_s(m,k) = 0$ because it is not possible to have k edges in less than $\sqrt{k}$ nodes. Hence we can rewrite the previous equation as:

$$\theta(m,k) = \sum_{s=1}^{\lfloor m-\sqrt{k} \rfloor} \theta_s(m,k) \tag{19}$$

Moreover θ_s can be rewritten as follows:

$$\theta_s(m,k) = \binom{m}{s} \sigma(m-s,k), \tag{20}$$

where $\sigma(m-s,k)$ is the number of graphs with $m-s$ nodes, k edges and no isolated node. This value is multiplied by the number of ways to choose the s isolated nodes within the m nodes, which is $\binom{m}{s}$. We have then:

$$\theta(m,k) = \sum_{s=1}^{\lfloor m-\sqrt{k} \rfloor} \binom{m}{s} \sigma(m-s,k) \tag{21}$$

Finally, σ the number of graphs without isolated nodes is equal to the total number of graphs minus the number of graphs with at least one isolated node, and so: $\sigma(m-s,k) = \binom{(m-s)^2}{k} - \theta(m-s,k)$. Leading to the result. ■

Note that Eq.(17) can be viewed as a non-homogeneous linear differential equation with variable coefficients. This type of equations can be written in a closed form [15].

The same reasoning can be used to solve the problem for a directed graph without self-loops or for an undirected graph with or without self-loops. In these cases the m^2-2 in the limit of the sum in Eq. (13) should be replaced respectively with $m(m-1)-2$, $m(m+1)/2-1$ and $m(m-1)/2-1$ and $\binom{m^2}{k}$ the total number of graph with m nodes and k edges should be replaced respectively with $\binom{m(m-1)}{k}$n $\binom{m(m+1)/2}{k}$ and $\binom{m(m-1)/2}{k}$.

Considering that there are *relatively* few graphs which are not connected while having no isolated nodes, the two probabilities $\bar{\rho}(m,p)$ and $\bar{\rho}_0(m,p)$ appears to be really close. It seems even that the two probabilities tend towards each other for large m. This leads to the following conjecture.

Conjecture 1: Let $\bar{\rho}(m,p)$ be the probability that a random (m,p)-Erdős-Rényi graph is connected and $\bar{\rho}_0(m,p)$ be the probability that a random (m,p)-Erdős-Rényi graph has no isolated nodes, then:

$$\lim_{m\to\infty} \frac{\bar{\rho}(m,p)}{\bar{\rho}_0(m,p)} = 1 \tag{22}$$

This result would straightforwardly lead to this second result:

$$\lim_{n\to\infty} \frac{\rho(n,p,M)}{\rho_0(n,p,M)} = 1 \tag{23}$$

where $\rho_0(n,p,M) = \prod_{i=1}^{\hat{n}} \bar{\rho}_0(m_i,p)$ is then the upper-bound on $\rho(n,p,M)$.

C. Simulations

To illustrate the result of Theorem 1 we compare an experimental value of $\bar{\rho}(m,p)$ and the theoretical value of $\bar{\rho}_0(m,p)$ given in Eq. (13). To find the experimental value we generate 1000 (m,p)-Erdős-Rényi graphs for each value of of $m = \{6,8,10,12\}$ and $p = \{0,0.02,0.04,\ldots,1\}$. An approximation of $\bar{\rho}(m,p)$ consists in counting the proportion of the 1000 graphs that are connected. The results are presented in Fig. 4. We notice that i) the inequation $\bar{\rho}_0 \geq \bar{\rho}$ is well verified and ii) the upper-bound $\bar{\rho}_0$ is really close to $\bar{\rho}$ and the difference seems to vanish when m grows. Let us note that the formula given in Theorem 1 does not allow to compute $\theta(m,k)$ for large value of m. Indeed the recursive call to the function θ is computationally heavy and the value of the binomial coefficient grows quickly.

To illustrate the result of Proposition 1 and showing that the passage to the partitioning problem works well, we present a second simulation. To compute the experimental value of $\rho(n,p,M)$, as before, we generate 1000 Erdős-Rényi with $n = 100$ nodes for each value of p varying between 0 and 1. For each graph we generate a random partition with the cluster sizes $m_1 = m_2 = \ldots = m_{20} = 5$. An approximation of ρ consists in counting the proportion of the 1000 graphs whose partition is connected. The results are presented in Fig. 5.

V. Conclusion

When looking at graph partitioning problems, it is clear that imposing a connectedness constraint deteriorates the quality of the solution. We presented here an approach based on probability and counting allowing to estimate *the price of connectedness* which is the degradation due to the connectedness constraint. This is an high-level approach in the sense that it does not focus on the particular problem, only on the cardinalities of the partition sets in the constraint and unconstrained cases. The drawback of this high-level approach is that, the results are not precise when considering particular problem. Then an extension of this work would be to investigate to what extent, the ratio of connectedness (i.e. the relative size of the research set in the constrained problem) influences the price of connectedness. In other words, the study of the function $\Delta(\rho)$ which depends on the particular problem considered. Another extension of this work would be the simplification of the recursive formula given in Theorem 1 allowing to make it computable for large-scale graphs.

References

[1] N. Monshizadeh, H. L. Trentelman, and M. K. Camlibel, "Stability and synchronization preserving model reduction of multi-agent systems," *Systems & Control Letters*, vol. 62, no. 1, pp. 1–10, 2013.

[2] X. Cheng, Y. Kawano, and J. M. Scherpen, "Graph structure-preserving model reduction of linear network systems," in *Control Conference (ECC), 2016 European*. IEEE, 2016, pp. 1970–1975.

[3] N. Martin, P. Frasca, and C. Canudas de Wit, "Large-scale network reduction towards scale-free structure," *IEEE Transactions on Network Science and Engineering*, pp. 1–12, 2018. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01885140

[4] T. Ishizaki, K. Kashima, J.-i. Imura, and K. Aihara, "Model reduction and clusterization of large-scale bidirectional networks." *IEEE Trans. Automat. Contr.*, vol. 59, no. 1, pp. 48–63, 2014.

[5] T. Ishizaki, K. Kashima, A. Girard, J.-I. Imura, L. Chen, and K. Aihara, "Clustered model reduction of positive directed networks," *Automatica*, vol. 59, pp. 238–247, 2015.

[6] M. O. Jackson, *Social and economic networks*. Princeton university press, 2010.

[7] T. Ishizaki, K. Kashima, A. Girard, J.-i. Imura, L. Chen, and K. Aihara, "Clustering-based $\mathscr{H}_2$-state aggregation of positive networks and its application to reduction of chemical master equations," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 4175–4180.

[8] T. Ishizaki, K. Kashima, J.-i. Imura, and K. Aihara, "Model reduction of multi-input dynamical networks based on clusterwise controllability," in *American Control Conference (ACC), 2012*. IEEE, 2012, pp. 2301–2306.

[9] P. Erdos and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, no. 1, pp. 17–60, 1960.

[10] S. Mussmann, "Counting partitions or subsets," https://www.projectrhea.org/rhea/index.php/Counting_subsets_of_sets#Partitions_Of_Sets_into_Fixed_Sizes, [Online; accessed October-2018].

[11] M. Boguná and M. Á. Serrano, "Generalized percolation in random directed networks," *Physical Review E*, vol. 72, no. 1, p. 016106, 2005.

[12] A. J. Graham and D. A. Pike, "A note on thresholds and connectivity in random directed graphs," *Atl. Electron. J. Math*, vol. 3, no. 1, pp. 1–5, 2008.

[13] J. Hopcroft, "Advanced design and analysis of algorithms - lecture 7: More on graph connectivity," Feb. 2008.

[14] N. Galli and K. Simon, "On the distribution of the number of isolated vertices in random undirected graphs," *Institute for Theoretical Computer Science, ETH Zurich*, vol. 11.

[15] R. K. Mallik, "Solutions of linear difference equations with variable coefficients," *Journal of mathematical analysis and applications*, vol. 222, no. 1, pp. 79–91, 1998.

Appendix

A. Algorithms

We describe here the algorithm proposed in [5] to solve the problem (3) and how we have adapted it to solve the constrained problem (6). The solutions brought by these algorithms are not the optimal solution of the minimisation problem but a good approximation of it.

The algorithm is an iterative greedy algorithm which means that at each step the best pair of nodes is chosen and merged. As at each step only two nodes are merged, the number of nodes in the graph is reduced by one, and then the same procedure is applied $(n-\hat{n})$ times where n is the size of the initial graph and $\hat{n}$ is the desired size of the

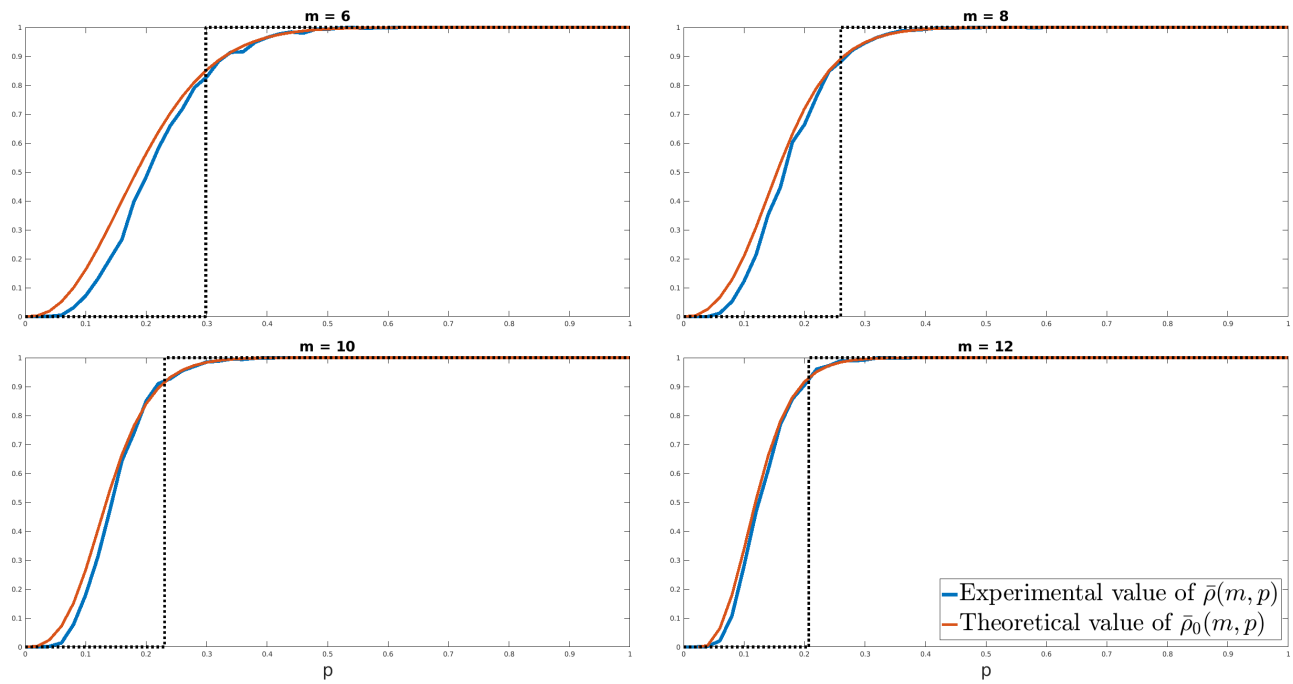


Fig. 4: Comparison between the experimental value of $\bar{\rho}(m,p)$ (the probability that a given (m,p)-Erdős-Rényi graph is connected) and the theoretical value of the upper-bound $\bar{\rho}_0(m,p)$ (the probability that a given (m,p)-Erdős-Rényi graph has no isolated node) for $p \in 0:0.02:1$ and for different values of m. The phase transition behaviour detailed in Eq. (12) is represented in dotted black: when m grows $\bar{\rho}(m,p)$ gets closer to the transition.

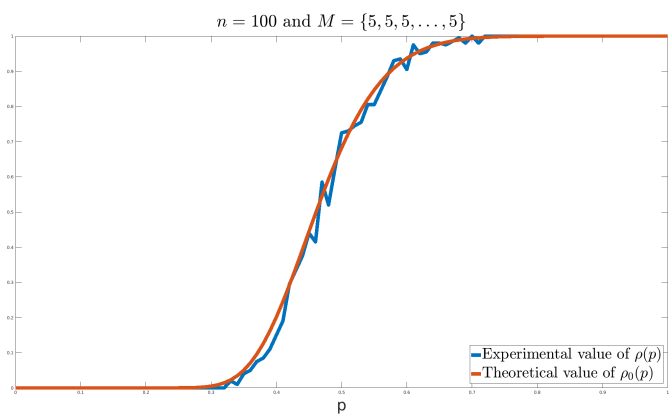


Fig. 5: Comparison between the experimental value of $\rho(100,p,\{5,\ldots,5\})$, and the theoretical value of the upper-bound $\rho_0(100,p,\{5,\ldots,5\})$ for $p \in 0:0.005:1$.

final graph. Both algorithms are described in 1 as there is only a small modification. In the algorithm solving the unconstrained problem the part *such that* $A(i,j) \neq 0$ at line 4 is removed. Indeed in the unconstrained problem we apply the same algorithm expect that two nodes can be merged without being connected (which is the removed condition).

Other computations are done after this step but there are identical for the two problems. For our concern we are only interested in this part of the reduction method.

Algorithm 1 Cluster Model Reduction

Input: Initial system $\Sigma : x(k+1) = Ax(k) + Bu(k)$ represented by the graph $G(A)$ (where A is adjacency matrix of G) and a desired final size m

Output: Final partition I

1: Initialise the partition as $I \leftarrow \{I_1,\ldots,I_N\} = \{\{1\},\ldots,\{N\}\}$

2: Solve the Lyapunov equation $A\Phi A^T - \Phi + BB^T = 0$

3: **for** $k = 1:(n-m)$ **do**

4: **for** all pair $i > j$ *such that* $A(i,j) \neq 0$ **do**

5: Assign the reducibility value to the pair (I_i, I_j) as:

$$\delta_{I_i,I_j} = \mathrm{Tr}(\Phi_{I_l \cup I_j}) - \frac{\mathbf{1}\Phi_{I_l \cup I_j}\mathbf{1}}{n}$$

6: **end for**

7: Merge the clusters minimising the reducibility value, which is if $(I_{i_0}, I_{j_0}) := \arg\min \ \delta(I_i, I_j)$ then $I_{i_0} \leftarrow I_{i_0} \cup I_{j_0}$ and $I \leftarrow I \backslash I_{j_0}$

8: **end for**